

Supporting research data services by LOD based authority data

Timo Borst

Workshop: Metadata and Persistent Identifiers
 for Social and Economic Data

Berlin

7.-8.5.2012

Overview

1. Metadata handling in current information systems
2. Linking metadata to authority data (persons)
3. Embedding authority data into library and information workflows
4. Summary and outlook

Metadata handling in current information systems

- Information systems
 - String based capture with (if at all) internal control of names, institutions, subjects etc.
- Library systems
 - Linking to internal identifiers (e.g. German PPN identifier), in a collaborative, but ‚closed‘ environment (intransparent and non-accessible to others)
 - Use of proprietary control characters and delimiters to „tag“ identifiers
 - No continuous policy for introducing authority files (no general guideline, but case based introduction)

Cataloguing and metadata maintenance

- Aggregation: In a distributed & networked environment, metadata is harvested and aggregated from different sources (not only library systems, but repositories, publishers' data, web pages etc.)
- Especially concerning research data, you cannot expect metadata to be located in one place (different registries for different disciplines and data providers)
- Lessons to be learned from harvesting publication metadata:
 - Heterogenous in spelling -> ambiguous names
 - Compliance in spelling
 - Significant amount of duplicates
 - Registries for research data (yet) do not help

Search (all) Internet Sources E

orszag

orszag, jonathan michael

orszag, peter r.

orszag, mike

orszag, j. michael

orszag, peter

orszag, j.michael

orszag, j.m.

orszag j

orszag, j michael

orszag, jonathan m.

0500 Asu
 1100 2008
 1500 /leng
 3010 Gupta, Vrinda
 3011 Jin, Henry!509457088!Jin, Henry Hongbo ; ID: gnd/131459201
 3012 Orszag, Michael!061374962!Orszag, Jonathan Michael ; ID: gnd/171198492
 3013 Piggott, John!39186923X!Piggott, John ; ID: gnd/129263117
 4000 How do Australian superannuation fund trustees perceive their role and effectiveness? / Vrinda Gupta, Henry Jin, Michael Orszag and John Piggott
 4061 graph. Darst.
 4070 /j2008/p200-219
 4241 In: !563482893!Pension fund governance. - Cheltenham [u.a.] : Elgar%200800000000000800%
 5249 //50*Aufsatz im Buch -- Article in book

LOK

4901 08-09-08; 7816/0206
 6110 k1

E01

7001 08-09-08 : z1k
 7901 08-09-08; 7816/0206
 7800 889887543

Cataloguing and metadata maintenance

Approach to tackle these issues (among others for disambiguation, e.g. algorithmic and social)

- Metadata fields for persons, subjects and organizations are linked to identifiers which are persistent, accessible and linkable according to LOD principles (stable URIs, HTTP-requests, content negotiation, statements about these identifiers etc.)
- Local customizations (language dependent labels for persons, subjects etc.) still must be possible
- In the future, registries for research data may handle author identification better than traditional biographical data

Data modelling based on Semantic Web / SKOS standards

- Modelling of concepts...

URI	lang_de	lang_en	lang_fr
http://zbw.eu/stw/descriptor/19664-4	Finanzmarkt-krise	financial crisis	crise financière
broader Terms	narrower Terms	altLabel	close Match
<stw/descriptor/10343-6> <stw/descriptor/18730-1>	<stw/descriptor/10343-6> <stw/descriptor/18730-1>	"Financial instability" @en "Finanzmarkt-krise"@de "Krise der Finanzmärkte"@de	http://dbpedia.org/resource/Financial_crisis

Table 1 Concept representation with URIs.

- ...and persons according to SemWeb principles and SKOS standard:

uri	surname	forename	variantNames	academicTitle	lifeData	affiliations
http://d-nb.info/gnd/124825109	"Snower"	"Dennis J."	Snower, Dennis James Snower, D. J. Snower, Dennis Snower, D.	Prof. Ph.D.	1950-	http://d-nb.info/gnd/1007681-5

Repräsentation

Example PND-/GND-Record

```
<gnd:DifferentiatedPerson rdf:about="http://d-nb.info/gnd/118598546">
  <gnd:preferredNameForThePerson>Benedikt &lt;XVI., Papst&gt;</gnd:preferredNameForThePerson>
  <gnd:preferredNameEntityForThePerson rdf:parseType="Resource">
    <gnd:epithetGenericNameTitleOrTerritory>Papst</gnd:epithetGenericNameTitleOrTerritory>
    <gnd:counting>XVI.</gnd:counting>
    <gnd:personalName>Benedikt</gnd:personalName>
  </gnd:preferredNameEntityForThePerson>
  <gnd:variantNameForThePerson>Ratzinger, Joseph</gnd:variantNameForThePerson>
  <gnd:variantNameEntityForThePerson rdf:parseType="Resource">
    <gnd:forename>Joseph</gnd:forename>
    <gnd:surname>Ratzinger</gnd:surname>
  </gnd:variantNameEntityForThePerson>
  <gnd:gender rdf:resource="http://d-nb.info/standards/vocab/gnd/Gender#male"/>
  <gnd:dateOfBirth>1927</gnd:dateOfBirth>
  <gnd:dateOfBirth>16.04.1927</gnd:dateOfBirth>
  <gnd:placeOfBirth rdf:resource="http://d-nb.info/gnd/4491052-6"/>
  <gnd:familialRelationship rdf:resource="http://d-nb.info/gnd/119141582"/>
  <gnd:biographicalOrHistoricalInformation>Sohn von Joseph und Maria Ratzinger, Papst seit
    19.04.2005</gnd:biographicalOrHistoricalInformation>
  <gnd:professionOrOccupation rdf:resource="http://d-nb.info/gnd/4059756-8"/>
  <gnd:professionOrOccupation rdf:resource="http://d-nb.info/gnd/4044561-6"/>
  <gnd:placeOfActivity rdf:resource="http://d-nb.info/gnd/4050471-2"/>
  <gnd:placeOfActivity rdf:resource="http://d-nb.info/gnd/4062404-3"/>
  <gnd:geographicAreaCode
    rdf:resource="http://d-nb.info/standards/vocab/gnd/geographic-area-code#XA-VA"/>
  <gnd:geographicAreaCode
    rdf:resource="http://d-nb.info/standards/vocab/gnd/geographic-area-code#XA-DE"/>
  <gnd:gndIdentifier>118598546</gnd:gndIdentifier>
  <gnd:oldAuthorityNumber>(DE-588c)310182379</gnd:oldAuthorityNumber>
</gnd:DifferentiatedPerson>
```

Additional context information provided by authority data

- preferred labels
 - alternate labels
 - gender
 - life data
 - profession and occupations
 - family relationships
 - (potential) links to other (external) information like
 - other identifier systems (VIAF, ORCID, INSI, PND, ...)
 - additional researcher information (from DBPedia or other LOD-datasets)
 - list of publications associated with the person
 - contact data (to be maintained by the researcher)
-

Additional context information provided by authority data

...all this data is supposed to be used for identifying and disambiguating persons!

Embedding authority data into library and information workflows

Example: da|ra (registration agency for social and economic data)

- Uploading metadata
 - Autosuggesting names with additional information for identification <http://zbw.eu/beta/econ-ws/examples/suggest2.html>
- Updating (and curating) metadata (from external sources)
 - Associating the name string with the identifier whose (preferred) labels (plus additional information) are exactly or closely matched.

Example: dataset <http://hdl.handle.net/1902.1/13772>, operation:
determining contributor P. Martin

Embedding authority data into library and information workflows

- Searching data
 - Searching for authors/contributors: Autosuggest of names associated with an authority record + research data
- Showing data
 - Once associated with identifiers, results could be augmented with additional linked data presented e.g. as a tooltip (to be solved: nested resources as URIs)

The screenshot shows a dataset entry for "Globalization and Emerging Markets: With or Without Crash? [Dataset]" by P. Martin and H. Rey. A mouse cursor is hovering over the author name "P. Martin", which triggers a tooltip. The tooltip contains detailed information about Philippe J. Martin, including his affiliations: Sciences-Po (Paris), CEPR, Federal Reserve Bank (New York, NY), Centre d'Études Prospectives et d'Informations Internationales (Paris), Université Panthéon-Sorbonne (Paris), and Center for Economic Policy Research (Stanford, Calif.). To the right of the dataset title, additional metadata is displayed: the handle "hdl:1902.1/13772", "53 downloads", and "Last Released: Mar 12, 2010".

Globalization and Emerging Markets: With or Without Crash? [Dataset]

by P. Martin; H. Rey

Produced by Sciences Po

Distributed by Dataaverse Network (DAN)

Related View Publication

hd1:1902.1/13772

53 downloads

Last Released: Mar 12, 2010

Martin, Philippe J. (Wirtschaftswissenschaftler; Tätig bei Sciences-Po (Paris) und CEPR; Federal Reserve Bank <New York, NY>, Centre d'Études Prospectives et d'Informations Internationales <Paris>, Université Panthéon-Sorbonne <Paris>, Center for Economic Policy Research <Stanford, Calif.>)

Start entering term:

martin

Sala-i-Martin, Xavier (1963 - ; Wirtschaftswissenschaftler; Tätig als Prof. für Wirtschaft an
Martin, Will (1953 - ; Wirtschaftswissenschaftler; International Bank for Reconstruction and
Martinez-Vazquez, Jorge (1948 - ; Wirtschaftswissenschaftler; Tätig an der Georgia State
Martin, Philippe J. (Wirtschaftswissenschaftler; Tätig bei Sciences-Po (Paris) und CEPR;
Martin, Albert (1949 - ; Prof. für Betriebswirtschaftslehre; Schwerpunkt: Personal und Füh
Martin, Stephen (1948 - ; Wirtschaftswissenschaftler; Michigan State University , Kranner
Martin, Philip L. (1949 - ; Wirtschaftswissenschaftler; University of California / Davis Camp
Martin, Ron (1948 - ; Prof. of Economic Geography; Brit. Geograph; University)
Martínez Pería, María Soledad (Wirtschaftswissenschaftlerin; Senior Economist in the Fir

Selected terms:

- [Martin, Philippe J.](#) 
- [Martin, Philip L.](#) 

Martin, Philippe J. (Wirtschaftswissenschaftler; Tätig bei Sciences-Po (Paris) und CEPR;
Federal Reserve Bank <New York, NY>, Centre d'Études Prospectives et d'Informations
Internationales <Paris>, Université Panthéon-Sorbonne <Paris>, Center for Economic
Policy Research <Stanford, Calif.>)

Start entering term:

1. 2. Title

Globalization and Emerging Markets: With or Without Crash? [Dataset]

1. 3. Other Titles

 Add Title

1. 4. Principal Investigators



Person

First Name

Philippe

Middle Name

J.

Last Name *

Martin

Affiliation

Centre d'Etudes Prospectives et d'Informations Internationales



Add Affiliation ID

Person ID

ID *

124546781

Vocabulary of Person ID *

gnd

URI Authority Dataset

http://d-nb.info/gnd/124546781



Add Person ID

Summary and outlook

- The system of scholarly communication is in general dependent on a web of trust, and in particular dependent on reliable identity management
- Information and library systems yet do not support this
- Libraries and information providers could establish a backbone for authority data which is crucial for identifying persons on the web
- Semantic Web technology could help to mesh up local information with a layer of centrally provided and updated authority data
- To become first-class objects like publications, research data resp. their records should be linked from authority data



Timo Borst
t.borst@zbw.eu

Identifier systems

- Which one to chose?

„It will take some time to determine which agent or coalition of agents in the scholarly communication game will end up providing the most reliable and useful author identification system. [J. Pitman (2012), Author Identity and Open Bibliography, <http://bulletin.imstat.org/2012/03/author-identity-and-open-bibliography/>]

- Criteria

- Coverage?
- Integration with other identifier systems?
- Availability?
- Machine readable formats?
- Support by publishers?
- Support by researchers (authorising)?
- Terms of use / Licensing?

- Services?

Systems for (author) identification

Candidates:

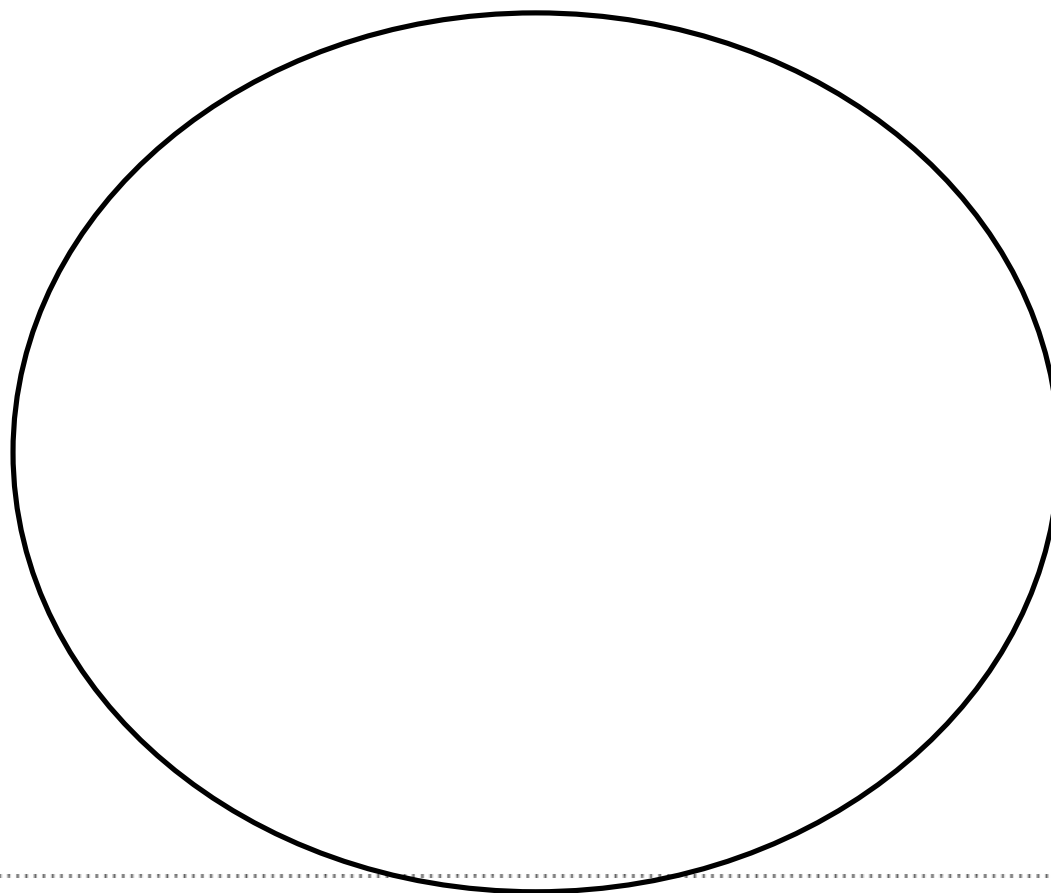
- VIAF (,Virtual Internet Authority File‘ <http://viaf.org/>)
 - Players: National Libraries, OCLC
 - Goal: Interlinking of national authority records
 - Status:
- ORCID
 - Players:
 - Goal:
 - Status:
- ISNI (,International Name Standard Identifier‘ <http://www.isni.org/>)
 - Players: Societies, Proquest, National Libraries, OCLC
 - Goal: To serve as a ,bridge identifier‘ for different local identifier systems
 - Status: ISO Standard, supported by OCLC

Provision of and access to Linked Authority Data

- Provision (RDF, RDFa, JSON, SPARQL-Endpoint, Turtle-files,...)
- Access: ReST (other approaches?)
- Lightweight integration into applications and augmentation of local, yet unconnected information

Authority data – other categories

According to German National Library (DNB):



Systems for (author) identification (II)

- PND
 - Players:
 - Goal:
 - Status:
- URI
 - Players:
 - Goal:
 - Status: